```
OM of: US-09-126-945B-2_COPY_236_243 to: GenEmbl:*   out_format : pfs

Date: Sep 7, 2001  6:28 PM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlh
 -Q=/cgn2_1/USPTO_spool/US09126945/runat_07092001_152108_3653/app_query.fasta_1.487
 -DB=GenEmbl -QFMT=fastap -SUFFIX=olig.rge -GAPOP=4.500
 -GAPEXT=0.050 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=60.000 -XGAPEXT=60.000
 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=60.000 -YGAPEXT=60.000
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=oligo
 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=quality
 -THR_MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
 -MAXLEN=2000000000 -USER=US09126945_@CGN1_1_7055 -NCPU=6
 -ICPU=3 -LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY
 -WAIT -THREADS=1

Search information block:
Query: US-09-126-945B-2_COPY_236_243
Query length: 8
Database: GenEmbl:*
Database sequences: 1344157
Database length: -856060004
Search time (sec): 1645.250000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP=XGAPOP=60.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT=XGAPEXT=60.000


score_list:
Sequence              Strd Orig    ZScore    EScore Len   ! Documentation  ..
gb_pat1:AX048148       +    8.00    166.55    0.4860   1894  ! AX048148 Sequence 1 from Pater
gb_pr4:AF071538        +    8.00    166.55    0.4860   1894  ! AF071538 Homo sapiens Ets trar
gb_pr1:AB031549        +    8.00    166.53    0.4869   1898  ! AB031549 Homo sapiens PSE mRNA
gb_pat1:AX048159       +    8.00    162.65    0.8014   3317  ! AX048159 Sequence 12 from Pate
gb_in2:CEC46C2         +    8.00    145.13    7.58     41115  ! Z68296 Caenorhabditis elegans
gb_pr6:AL157372        -    8.00    136.09    24.15    150660  ! AL157372 Human DNA sequence
gb_ba2:AF320001        +    7.00    134.86    28.30    7698  ! AF320001 Acinetobacter sp. BD4
gb_ba3:MBDR3S1         -    7.00    133.55    33.44    9281  ! U35017 Mycobacterium bovis del
gb_pat2:I86264         -    7.00    131.53    43.35    12412  ! I86264 Sequence 18 from paten
gb_ba3:MTCY336         +    7.00    124.85    102.17   32437  ! Z95586 Mycobacterium tubercul
gb_pl1:AB026650        -    7.00    120.06    188.87   64569  ! AB026650 Arabidopsis thaliana
gb_pr2:AC008530        +    7.00    117.81    251.75   89095  ! AC008530 Homo sapiens chromos
gb_htg24:AP002018      +    7.00    115.42    342.22   125674  ! AP002018 Homo sapiens chromc
gb_htg5:AC016212       -    7.00    113.73    425.04   160217  ! AC016212 Homo sapiens clone
gb_pr2:AC005747        +    7.00    113.45    440.36   166701  ! AC005747 Homo sapiens chromc
gb_htg9:AC024629       -    7.00    113.33    447.31   169650  ! AC024629 Homo sapiens chromc
gb_htg19:AC091071      -    7.00    113.15    458.05   174218  ! AC091071 Oryza sativa chromc
gb_htg24:AP003416      -    7.00    113.04    464.53   176982  ! AP003416 Oryza sativa chromc
gb_htg20:AL158154      -    7.00    112.96    469.49   179104  ! AL158154 Homo sapiens chromc
gb_htg16:AC079219      +    7.00    112.88    473.95   181008  ! AC079219 Mus musculus chromc
gb_htg20:AL353773      +    7.00    112.85    476.10   181931  ! AL353773 Homo sapiens chromc
gb_pr5:AL133417        -    7.00    112.76    481.43   184212  ! AL133417 Human DNA sequence
gb_htg19:AC091088      -    7.00    112.71    484.39   185481  ! AC091088 Oryza sativa chromc
gb_htg24:AP003436      +    7.00    112.69    485.52   185967  ! AP003436 Oryza sativa chromc
gb_htg20:AL355380      -    7.00    112.68    486.30   186302  ! AL355380 Homo sapiens chromc
gb_pr2:AC006239        +    7.00    112.36    506.59   195032  ! AC006239 Homo sapiens chromc
gb_htg14:AC067751      +    7.00    111.44    569.89   222537  ! AC067751 Homo sapiens chromc
gb_htg12:AC044804      +    7.00    110.55    639.31   253123  ! AC044804 Mus musculus chromc
gb_htg20:AL353740      -    7.00    110.34    656.40   260714  ! AL353740 Homo sapiens chromc
gb_pr5:AF213884S2      +    7.00    110.18    670.37   266941  ! AF224669 Homo sapiens mannos
gb_sts2:G23557         +    6.00    134.18    30.86    363  ! G23557 human STS WI-30916, sequ
gb_ov:CMU23969         -    6.00    131.17    45.37    559  ! U23969 Carpodacus mexicanus isc
gb_ov:CMU23968         -    6.00    130.66    48.48    602  ! U23968 Carpodacus mexicanus isc
gb_in2:AF259381        -    6.00    129.95    53.12    667  ! AF259381 Crisia denticulata Cd8
gb_ov:P0UMTCBF         +    6.00    129.30    57.72    732  ! D50025 Saw shark Mitochondrial
gb_pl3:CNS01DED        -    6.00    128.86    61.09    780  ! AL116717 Botrytis cinerea strai
gb_vi1:AF004437        +    6.00    128.62    62.97    807  ! AF004437 Venezuelan equine ence
gb_sts1:CNS01LH9       +    6.00    128.57    63.39    813  ! AL149630 Anopheles gambiae STS
gb_vi1:AF004451        +    6.00    128.56    63.46    814  ! AF004451 Mucambo virus strai
gb_vi1:AF004454        +    6.00    128.56    63.46    814  ! AF004454 Tonate virus strain
gb_vi1:AF004462        +    6.00    128.56    63.46    814  ! AF004462 Bijou Bridge virus
gb_vi2:VEU55343        +    6.00    128.55    63.53    815  ! U55343 Venezuelan equine enc
gb_vi2:VEU55344        +    6.00    128.55    63.53    815  ! U55344 Venezuelan equine enc
gb_vi2:VEU55345        +    6.00    128.55    63.53    815  ! U55345 Venezuelan equine enc
gb_vi2:VEU55348        +    6.00    128.55    63.53    815  ! U55348 Venezuelan equine enc

seq_name: gb_pat1:AX048148

seq_documentation_block:
LOCUS       AX048148     1894 bp    DNA              PAT       15-DEC-2000
DEFINITION  Sequence 1 from Patent WO0070092.
ACCESSION   AX048148
VERSION     AX048148.1  GI:11876967
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1894)
  AUTHORS   Kaufmann,J. and Xin,H.
  TITLE     Expression of ets-domain proteins in cancer
  JOURNAL   Patent: WO 0070092-A 1 23-NOV-2000;
            Chiron Corporation (US)
FEATURES             Location/Qualifiers
     source          1. .1894
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     CDS             416. .1423
                     /note="Human JKETS."
                     /codon_start=1
                     /protein_id="CAC19002.1"
                     /db_xref="GI:11876968"
                     /translation="MGSASPGLSSVSPSHLLLPPDTVSRTGLEKAAAGAVGLERRDWS
                     PSPPATPEQGLSAFYLSYFDMLYPEDSSWAAKAPGASSREEPPEEPEQCPVIDSQAPA
                     GSLDLVPGGLTLEEHSLEQVQSMVVGEVLKDIETACKLLNITADPMDWSPSNVQKWLL
                     WTEHQYRLPPMGKAFQELAGKELCAMSEEQFRQRSPLGGDVLHAHLDIWKSAAWMKER
                     TSPGAIHYCASTSEESWTDSEVDSSCSGQPIHLWQFLKELLLKPHSYGRFIRWLNKEK
                     GIFKIEDSAQVARLWGIRKNRPAMNYDKLSRSIRQYYKKGIIRKPDISQRLVYQFVHP
                     I"
BASE COUNT      368 a    653 c    571 g    302 t
ORIGIN


alignment_scores:
            Quality:    8.00              Length:      8
              Ratio:   1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_236_243 x AX048148   ..

 Align seg 1/1  to: AX048148  from: 1  to: 1894

       1 ThrAspSerGluValAspSerSer 8
         ||||||||||||||||||||||||
    1121 ACCGACAGCGAGGTGGACTCATCA 1144

seq_name: gb_pr4:AF071538

seq_documentation_block:
LOCUS       AF071538     1894 bp    mRNA             PRI       10-JAN-2000
DEFINITION  Homo sapiens Ets transcription factor PDEF (PDEF) mRNA, complete
            cds.
ACCESSION   AF071538
VERSION     AF071538.1  GI:4007417
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1894)
```

```
  AUTHORS   Oettgen,P., Finger,E., Sun,Z., Akbarali,Y., Thamrongsak,U.,
            Boltax,J., Grall,F., Dube,A., Weiss,A., Brown,L., Quinn,G., Kas,K.,
            Endress,G., Kunsch,C. and Libermann,T.A.
  TITLE     PDEF, a novel prostate epithelium-specific ets transcription
            factor, interacts with the androgen receptor and activates
            prostate-specific antigen gene expression
  JOURNAL   J. Biol. Chem. 275 (2), 1216-1225 (2000)
  MEDLINE   20092894
REFERENCE   2  (bases 1 to 1894)
  AUTHORS   Libermann,T.A., Oettgen,P., Kunsch,C. and Finger,E.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUN-1998) Medicine, Beth Israel Deaconess Medical
            Center, 330 Brookline Ave., Boston, MA 02215, USA
FEATURES             Location/Qualifiers
     source          1. .1894
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /map="6p21.3"
                     /tissue_type="prostate"
     gene            1. .1894
                     /gene="PDEF"
     CDS             416. .1423
                     /gene="PDEF"
                     /note="prostate epithelium-specific"
                     /codon_start=1
                     /product="Ets transcription factor PDEF"
                     /protein_id="AAC95296.1"
                     /db_xref="GI:4007418"
                     /translation="MGSASPGLSSVSPSHLLLPPDTVSRTGLEKAAAGAVGLERRDWS
                     PSPPATPEQGLSAFYLSYFDMLYPEDSSWAAKAPGASSREEPPEEPEQCPVIDSQAPA
                     GSLDLVPGGLTLEEHSLEQVQSMVVGEVLKDIETACKLLNITADPMDWSPSNVQKWLL
                     WTEHQYRLPPMGKAFQELAGKELCAMSEEQFRQRSPLGGDVLHAHLDIWKSAAWMKER
                     TSPGAIHYCASTSEESWTDSEVDSSCSGQPIHLWQFLKELLLKPHSYGRFIRWLNKEK
                     GIFKIEDSAQVARLWGIRKNRPAMNYDKLSRSIRQYYKKGIIRKPDISQRLVYQFVHP
                     I"
BASE COUNT      368 a    653 c    571 g    302 t
ORIGIN


alignment_scores:
            Quality:    8.00             Length:      8
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_236_243 x AF071538   ..

 Align seg 1/1  to: AF071538  from: 1  to: 1894

       1 ThrAspSerGluValAspSerSer 8
         ||||||||||||||||||||||||
    1121 ACCGACAGCGAGGTGGACTCATCA 1144

seq_name: gb_pr1:AB031549

seq_documentation_block:
LOCUS       AB031549     1898 bp    mRNA            PRI       20-JAN-2000
DEFINITION  Homo sapiens PSE mRNA for prostate ets, complete cds.
ACCESSION   AB031549
VERSION     AB031549.1  GI:6721497
KEYWORDS    prostate ets.
SOURCE      Homo sapiens adult male prostate epithelium prostate cancer
            cell_line:PC-3 cDNA to mRNA.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Yamada,N., Tamai,Y., Miyamoto,H. and Nozaki,M.
  TITLE     Cloning and expression of the mouse Pse gene encoding a novel Ets
            family member
  JOURNAL   Gene 241 (2), 267-274 (2000)
  MEDLINE   20137504

REFERENCE   2  (bases 1 to 1898)
  AUTHORS   Nozaki,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-AUG-1999) to the DDBJ/EMBL/GenBank databases. Masami
            Nozaki, Osaka University, Institute for Microbial Diseases; 3-1,
            Yamadaoka, Suita, Osaka 565-0871, Japan
            (E-mail:mnozaki@biken.osaka-u.ac.jp, Tel:81-6-879-8338,
            Fax:81-6-879-8339)
FEATURES             Location/Qualifiers
     source          1. .1898
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /cell_line="PC-3"
                     /cell_type="prostate cancer"
                     /chromosome="6"
                     /dev_stage="adult"
                     /map="6p21.2-21.3"
                     /sex="male"
                     /tissue_type="prostate epithelium"
                     /note="vector:pUC18"
     gene            406. .1413
                     /gene="PSE"
     CDS             406. .1413
                     /gene="PSE"
                     /codon_start=1
                     /product="prostate ets"
                     /protein_id="BAA89543.1"
                     /db_xref="GI:6721498"
                     /translation="MGSASPGLSSVSPSHLLLPPDTVSRTGLEKAAAGAVGLERRDWS
                     PSPPATPEQGLSAFYLSYFDMLYPEDSSWAAKAPGASSREEPPEEPEQCPVIDSQAPA
                     GSLDLVPGGLTLEEHSLEQVQSMVVGEVLKDIETACKLLNITADPMDWSPSNVQKWLL
                     WTEHQYRLPPMGKAFQELAGKELCAMSEEQFRQRSPLGGDVLHAHLDIWKSAAWMKER
                     TSPGAIHYCASTSEESWTDSEVDSSCSGQPIHLWQFLKELLLKPHSYGRFIRWLNKEK
                     GIFKIEDSAQVARLWGIRKNRPAMNYDKLSRSIRQYYKKGIIRKPDISQRLVYQFVHP
                     I"
BASE COUNT      380 a    651 c    567 g    300 t
ORIGIN


alignment_scores:
            Quality:    8.00             Length:      8
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_236_243 x AB031549   ..

 Align seg 1/1  to: AB031549  from: 1  to: 1898

       1 ThrAspSerGluValAspSerSer 8
         ||||||||||||||||||||||||
    1111 ACCGACAGCGAGGTGGACTCATCA 1134

seq_name: gb_pat1:AX048159

seq_documentation_block:
LOCUS       AX048159     3317 bp    DNA             PAT       15-DEC-2000
DEFINITION  Sequence 12 from Patent WO0070092.
ACCESSION   AX048159
VERSION     AX048159.1  GI:11876975
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 3317)
  AUTHORS   Kaufmann,J. and Xin,H.
  TITLE     Expression of ets-domain proteins in cancer
  JOURNAL   Patent: WO 0070092-A 12 23-NOV-2000;
            Chiron Corporation (US)
FEATURES             Location/Qualifiers
     source          1. .3317
                     /organism="Homo sapiens"
```

```
                     /db_xref="taxon:9606"
                     /note="GSEF-encoding sequence with promoter (Figs. 2A-2B)"
     promoter        1. .1380
                     /note="Promoter"
     CDS             1815. .2822
                     /note="GSEF coding sequence"
                     /codon_start=1
                     /protein_id="CAC19004.1"
                     /db_xref="GI:11876976"
                     /translation="MGSASPGLSSVSPSHLLLPPDTVSRTGLEKAAAGAVGLERRDWS
                     PSPPATPEQGLSAFYLSYFDMLYPEDSSWAAKAPGASSREEPPEEPEQCPVIDSQAPA
                     GSLDLVPGGLTLEEHSLEQVQSMVVGEVLKDIETACKLLNITADPMDWSPSNVQKWLL
                     WTEHQYRLPPMGKAFQELAGKELCAMSEEQFRQRSPLGGDVLHAHLDIWKSAAWMKER
                     TSPGAIHYCASTSEESWTDSEVDSSCSGQPIHLWQFLKELLLKPHSYGRFIRWLNKEK
                     GIFKIEDSAQVARLWGIRKNRPAMNYDKLSRSIRQYYKKGIIRKPDISQRLVYQFVHP
                     I"
BASE COUNT      710 a   1026 c    970 g    611 t
ORIGIN


alignment_scores:
            Quality:    8.00              Length:      8
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_236_243 x AX048159   ..

 Align seg 1/1  to: AX048159  from: 1  to: 3317

       1 ThrAspSerGluValAspSerSer 8
         ||||||||||||||||||||||||
    2520 ACCGACAGCGAGGTGGACTCATCA 2543

seq_name: gb_in2:CEC46C2

seq_documentation_block:
LOCUS       CEC46C2     41115 bp    DNA             INV       25-OCT-2000
DEFINITION  Caenorhabditis elegans cosmid C46C2, complete sequence.
ACCESSION   Z68296
VERSION     Z68296.1  GI:1130612
KEYWORDS    HTG; Serine/threonine-protein kinase; Transport protein.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
            Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 41115)
  AUTHORS   none.
  TITLE     Genome sequence of the nematode C. elegans: a platform for
            investigating biology. The C. elegans Sequencing Consortium
  JOURNAL   Science 282 (5396), 2012-2018 (1998)
  MEDLINE   99069613
  REMARK    The C.elegans Sequencing Consortium.
            Erratum:[[published errata appear in Science 1999 Jan
            1;283(5398):35 and 1999 Mar 26;283(5410):2103 and 1999 Sep
            3;285(5433):1493]]
REFERENCE   2  (bases 1 to 41115)
  AUTHORS   McMurray,A.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-DEC-1995) Nematode Sequencing Project, Sanger Centre,
            Hinxton, Cambridge CB10 1RQ, England and Department of Genetics,
            Washington University, St. Louis, MO 63110, USA. E-mail:
            jes@sanger.ac.uk or rw@nematode.wustl.edu
COMMENT     Coding sequences below are predicted from computer analysis, using
            predictions from Genefinder (P. Green, U. Washington), and other
            available information.
            Current sequence finishing criteria for the C. elegans genome
            sequencing consortium are that all bases are either sequenced
            unambiguously on both strands, or on a single strand with both a
            dye primer and dye terminator reaction, from distinct subclones.
            Exceptions are indicated by an explicit note.
            IMPORTANT: This sequence is not the entire insert of clone C46C2.
            It may be shorter because we only sequence overlapping sections
            once, or longer because we arrange for a small overlap between
            neighbouring submissions.
            The true left end of clone C46C2 is at 1 in this sequence. The true
            right end of clone C46C2 is at 104 in
            sequence AL032632.
            The start of this sequence (1. .104) overlaps with the end of
            sequence Z68295.
            The end of this sequence (41012. .41115) overlaps with the start of
            sequence AL032632.
            For a graphical representation of this sequence and its analysis
            see:- http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?
            name=C46C2.
FEATURES                Location/Qualifiers
     source             1. .41115
                        /organism="Caenorhabditis elegans"
                        /db_xref="taxon:6239"
                        /chromosome="IV"
                        /clone="C46C2"
     gene               join(13. .54,256. .446,589. .707,870. .1114,2461. .2648,
                        2993. .3485,3538. .3695,3830. .3941,4023. .4319)
                        /gene="C46C2.4"
     CDS                join(13. .54,256. .446,589. .707,870. .1114,2461. .2648,
                        2993. .3485,3538. .3695,3830. .3941,4023. .4319)
                        /gene="C46C2.4"
                        /note="contains similarity to Pfam domain: PF00561
                        (alpha/beta hydrolase fold), Score=30.8, E-value=1e-05,
                        N=1"
                        /codon_start=1
                        /protein_id="CAA92590.1"
                        /db_xref="GI:3874924"
                        /db_xref="SPTREMBL:Q18656"
                        /translation="MCSICSNSYLNEILQRYLYSHSFSLNNKKIALLYVSGQNTFNEN
                        ILKQGPFVQAAEEFGASMFALEHRFYGNSKPRSMNLTSKDLRYLKSSEAVQDIISFIN
                        YSNKKFNMNPGVKRSKYNEQVENSILKFGGQECFQKIKTGFNYLRAAMNTIQGRKELS
                        DQFKLEPRLNTQNPNSNEIQLFYLYVIAPFQQIVQFHRQQRFFKNSKYAKEGGPNFLC
                        IGQEGREDPNSIRFDVFAVVEKAQKFGATVYVLEHRFYGDSNVGDNSDLSKLSSLQML
                        YDLAEIIKEENLKTNTSNPWITFGGSYSGMLSAWMREIFHEFVVGAVASSAPILAKTD
                        FYEYIMVVEDVFRRYDIGCYNAIKKGFLEIQKMFLTEDGRDKLSKLFPSYPALRNNFS
                        ETRKHEFLLDLADPFETSVQYAGYGSGAFAKTSKKLKNKPKGDVDNQNDDISTENDGL
                        LWMWQTCTEFGFYQSTDTGNSIFGNVPVSYFVQQCMDLFGNNYTRATIDKQVGRTNHK
                        YDGTYEFNATNVVFLNGDADPWSPLGLKNSTDPSVVSFLINGTSHCVDMYSETEDDLP
                        DLKTARKIVDENIEKWLNSTTTTKSPNSANYPRALNSLAVLIVYYVIL"
     gene               complement(join(4878. .5195,5289. .5492,5628. .5819))
                        /gene="C46C2.5"
     CDS                complement(join(4878. .5195,5289. .5492,5628. .5819))
                        /gene="C46C2.5"
                        /note="cDNA EST yk579a1.3 comes from this gene
                        cDNA EST yk579a1.5 comes from this gene"
                        /codon_start=1
                        /protein_id="CAB54197.1"
                        /db_xref="GI:5824407"
                        /db_xref="SPTREMBL:Q9U3M0"
                        /translation="MASIGKLTDNQEMLKIFLIFTVLINCTVSVPADIPSLFSGLLNG
                        LSKNCLTETYSFLTELREIETLQNSKFNNVDVLYDKDLMKSLFGDLKSISLSIKSSNC
                        SSMIFAKYLLDGTVFIGDNIYGEAFNCLSKKRILVQIKYNSELQENLESLFMSCAREI
                        ILNNQSIDILEHLKPVSRCVAGKLACSKIDKNSFIGAVFAGVDLFDTFSNSRSILSQI
                        KENKMKLEFNKEKYYSYLN"
     gene               join(14510. .14554,15442. .15520,15825. .15894,16489. .16573,
                        16911. .17347,17403. .17704,17774. .18099,18147. .19391,
                        19680. .20378,20430. .21194,21281. .21712,22290. .22772,
                        23560. .23612,23668. .23872,24447. .24776)
                        /gene="C46C2.1"
     CDS                join(14510. .14554,15442. .15520,15825. .15894,16489. .16573,
                        16911. .17347,17403. .17704,17774. .18099,18147. .19391,
                        19680. .20378,20430. .21194,21281. .21712,22290. .22772,
                        23560. .23612,23668. .23872,24447. .24776)
                        /gene="C46C2.1"
                        /note="Similarity to Mouse A-RAF proto-oncogene
                        serine/threonine-protein kinase (SW:KRAA_MOUSE), contains
                        similarity to Pfam domain: PF00069 (Eukaryotic protein
                        kinase domain), Score=216.0, E-value=1.8e-61, N=1
                        cDNA EST yk15c4.3 comes from this gene; cDNA EST
                        EMBL:T01018 comes from this gene
```

cDNA EST yk28c9.3 comes from this gene; cDNA EST yk32b2.3 comes from this gene
cDNA EST yk15c4.5 comes from this gene; cDNA EST yk25h10.5 comes from this gene
cDNA EST yk28c9.5 comes from this gene; cDNA EST yk88d7.3 comes from this gene
cDNA EST yk68g2.3 comes from this gene; cDNA EST yk87e12.5 comes from this gene
cDNA EST yk88d7.5 comes from this gene; cDNA EST yk68g2.5 comes from this gene
cDNA EST yk161e2.3 comes from this gene; cDNA EST yk161e2.5 comes from this gene
cDNA EST yk169f8.3 comes from this gene; cDNA EST yk169f8.5 comes from this gene
cDNA EST yk439h9.3 comes from this gene; cDNA EST yk439h9.5 comes from this gene
cDNA EST yk431f12.3 comes from this gene; cDNA EST yk431f12.5 comes from this gene
cDNA EST yk419h3.3 comes from this gene; cDNA EST yk419h3.5 comes from this gene
cDNA EST yk348h7.3 comes from this gene; cDNA EST yk348h7.5 comes from this gene
cDNA EST yk348e7.3 comes from this gene; cDNA EST yk348e7.5 comes from this gene
cDNA EST yk324a11.3 comes from this gene; cDNA EST yk324a11.5 comes from this gene
cDNA EST yk309a2.3 comes from this gene; cDNA EST yk309a2.5 comes from this gene
cDNA EST yk323b11.3 comes from this gene; cDNA EST yk323b11.5 comes from this gene
cDNA EST yk315d9.3 comes from this gene; cDNA EST yk315d9.5 comes from this gene
cDNA EST yk288a8.3 comes from this gene; cDNA EST yk288a8.5 comes from this gene
cDNA EST yk212d9.3 comes from this gene; cDNA EST yk220a10.3 comes from this gene
cDNA EST yk220a10.5 comes from this gene; cDNA EST yk349g2.3 comes from this gene
cDNA EST yk214b1.5 comes from this gene; cDNA EST yk226f1.5 comes from this gene
cDNA EST yk345a8.5 comes from this gene; cDNA EST yk400f7.5 comes from this gene
cDNA EST yk405h3.5 comes from this gene; cDNA EST yk447d12.5 comes from this gene
cDNA EST yk552h2.3 comes from this gene; cDNA EST yk574a5.3 comes from this gene
cDNA EST yk552h2.5 comes from this gene; cDNA EST yk571g6.5 comes from this gene
cDNA EST yk574a5.5 comes from this gene"
/codon_start=1
/protein_id="CAA92591.1"
/db_xref="GI:3874925"
/db_xref="SPTREMBL:Q18657"
/translation="MTVICNRCRNISCRMSITNGGRPPAPPSSVSSTTASTTGNFGTR
RRLVNRIKKVDELHPAQENPTMGSHWLSEQERSRLEAVQQDWRRTRPMKWTSKKRPDD
PTTSSPSTVSISNALENSTPSLNNVSSITNSSSPFSLSSAATSTASAIIPFTSNVATN
HPHLNHHVSRIPQAIVTGGTNGSLPPLLISPTSAAAATPLISGKAGPMSPSTGSPINV
AATVLQNAVSSPQHSIFDRSRLNKIPPNTSLASSSSPSDAANNDKPIQQRHSILSNVR
TLTQAMVNDGPRTLTGDDMDKMVSEEERARKEQEKREEEEKAARRIDVEDDFDAQEKP
IDKSKNGRFLKFDEELGRGSFKTVFRGLDTETGVAVAWCELQESKLNKTERQRFREEA
EMLKDLQHPNIVRFYDYWESADLCGKRKYIVLVTELMTSGTLKMYLKRFKRINIKVVL
KSWCRQILKGLSFLHTRNPPVIHRDLKCDNIFITGTTGSVKIGDLGLATLKNKSFAKS
VIGTPEFMAPEMYEEMYDESVDVYAFGMCLLEMVTGEYPYSECMNPATIYRKVISGVK
PECFSRIPAQYPEIREIIDRCIRVRREERSTVKQLLVDDFFTPEDLIGIRVEIKNRDA
DLNDLNVEIQMQLRVYDEKKRKQYRFKENEGLQFAFDIENDSPDEVVQQMIEQQHIPD
EDTRMITKLIKDKVDAFRRDRDHRLLEIKRAKEEEERIREEAEIKEELRLRAEAKEKE
KERLEKERLEKKAAAAAAANPNPTPIPPTPATPHSSAQQQPIPPPLSTQTSAEIQQSA
QQPSVPVTMIANIPAMSPTSAQPQPVLSPTSAAVPVPTTMIHVPKPSEIPVQNVATTA
APVAANNVPPSPAPFKTEDIQTPTLAQNTVPRTISTDASGLVINTPASIASPSPAPSA
TDVASTTAPVTPAPTPTTTTDGGAAAASTTTENKEEKRKSNKRKVVMEILGCDESRNF
ALVSCRLDTSHKSVTFQFAPGTDKPCTIATKLLAEDCLLKVHVHIVEAQLGEVIQLIN
SDGKKGVGTKLATVLDPNSTEPPTITAVMPKDSSAATASNTKPKIEIEKTPPTRDASQ
EPNNVQVTNVRKVSQESNAESVQSIPRPGGIIVMSPTNQTDSAPPPTGAAAKPSRFQV
TKSADPIATPISSSISTATVIPIVAATPTNITSEPVIVQPITAQVITHLATPSPVSHS
LSSNSSPSATTHSNMSSIQSTTSVPGRRFTVQPVSQAESGISSSISTPHPEPTPAITS
CPPPVPSVPPVVSNGTLNLEVAPKQTPSATNQNVDTQHSSSTASTATLVSETPATVHV
TPISVPAPVQEPLVIDHHSDVLTQLDSELRKVSGVSHSASPSTVVESLTSMTPQTIPL
ACQTVPASIGQAPAVIAAAHAASLIPNASVPQSPSRLDAETGLAGLHEKLEALKMEQD
RREDMGDDAIGTTTTDGKDEIPIDTLKGLAEALGKVIHADGRETTPMPPDHPDLTDAS
TQQLISPSNPDVLTTMSSAVEGSASSTMIEDIDASTSAVDASMMNSMPPGAQNSTDQI
PAAMTLSMDQECAQSMTSSITRNTTGTKLATFENLETALSSTLGTHIRQPNAPSSRDE
TTAPMTPSFTNERIGGGGGGGATSFSIGTPPSHSPFPVSECDYDLKGQMDLESEDPEV
IQMIVRHRMEQHKLLEKQRVEIERLRSKIRVPRATSVNPEMIGDDEADTTLTALQSAL
GNASLSLPASPPPNTETTKVNTTVIPSDVLATRMTMSQSSTKSSNVSVSSRHRDNQSA
PPRHHHHQPHPPHHPHLQNHYHPPQNHTSATAPCPSAMVQLQAVSNNNVNPLHQPPHP
VSSQIPPQA"
gene join(30950. .31094,31314. .31774,31826. .32042,32101. .32212, 32259. .32364,32727. .33020,33104. .33258,33306. .33369)
/gene="C46C2.2"
CDS join(30950. .31094,31314. .31774,31826. .32042,32101. .32212, 32259. .32364,32727. .33020,33104. .33258,33306. .33369)

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-126-945B-2_COPY_236_243 x CEC46C2 ..

Align seg 1/1 to: CEC46C2 from: 1 to: 41115

```
      1 ThrAspSerGluValAspSerSer 8
        ||||||||||||||||||||||||
  38400 ACTGACTCGGAAGTGGACAGTTCA 38423
```

seq_name: gb_pr6:AL157372

seq_documentation_block:
LOCUS AL157372 150660 bp DNA PRI 29-NOV-2000
DEFINITION Human DNA sequence from clone RP11-375E1 on chromosome 6, complete sequence.
ACCESSION AL157372
VERSION AL157372.18 GI:11493236
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 150660)
AUTHORS Skuce,C.
TITLE Direct Submission
JOURNAL Submitted (29-NOV-2000) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT On Nov 30, 2000 this sequence version replaced gi:11342772.
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated repeat sequence elements. Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.
The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases:
Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping Group. Further information can be found at http://www.sanger.ac.uk/HGP/Chr6

```
          RP11-375E1 is from the library RPCI-11.2 constructed at the Roswell
          Park Cancer Institute by the group of Pieter de Jong. For further
          details see http://bacpac.med.buffalo.edu/
          VECTOR: pBACe3.6
          IMPORTANT: This sequence is not the entire insert of clone
          RP11-375E1 It may be shorter because we sequence overlapping
          sections only once, except for a 100 base overlap.
          The true left end of clone RP11-375E1 is at 1 in this sequence. The
          true left end of clone RP3-391O22 is at 150561 in this sequence.
          The true right end of clone RP11-302L10 is at 113044 in this
          sequence.
FEATURES             Location/Qualifiers
     source          1. .150660
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /clone="RP11-375E1"
                     /clone_lib="RPCI-11.2"
BASE COUNT    35470 a  38887 c  38740 g  37563 t
ORIGIN


alignment_scores:
           Quality:    8.00            Length:      8
             Ratio:  1.000              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_236_243 x AL157372/rev   ..

 Align seg 1/1  to reverse of: AL157372  from: 1  to: 150660

       1 ThrAspSerGluValAspSerSer 8
         ||||||||||||||||||||||||
  124658 ACCGACAGCGAGGTGGACTCATCA 124635

seq_name: gb_ba2:AF320001

seq_documentation_block:
LOCUS        AF320001     7698 bp    DNA             BCT       27-NOV-2000
DEFINITION   Acinetobacter sp. BD413 putative integral membrane protein,
             putative ATP binding protein of ABC transporter, competence factor
             ComA (comA), putative HtrB protein, and putative protease genes,
             complete cds; and unknown gene.
ACCESSION    AF320001
VERSION      AF320001.1  GI:11345436
KEYWORDS     .
SOURCE       Acinetobacter sp. BD413.
  ORGANISM   Acinetobacter sp. BD413
             Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae;
             Acinetobacter.
REFERENCE    1  (bases 1 to 7698)
  AUTHORS    Friedrich,A., Hartsch,T. and Averhoff,B.
  TITLE      Natural transformation in mesophilic and thermophilic bacteria:
             Identification and characterization of closely related ComA-like
             competence factors in Acinetobacter sp. BD413 and Thermus
             thermophilus HB27
  JOURNAL    Unpublished
REFERENCE    2  (bases 1 to 7698)
  AUTHORS    Friedrich,A., Hartsch,T. and Averhoff,B.
  TITLE      Direct Submission
  JOURNAL    Submitted (08-NOV-2000) Institute of Microbiology and Genetics,
             Universitaet Goettingen, Grisebachstrasse 8, Goettingen 37077,
             Germany
FEATURES             Location/Qualifiers
     source          1. .7698
                     /organism="Acinetobacter sp. BD413"
                     /strain="BD413"
                     /db_xref="taxon:104611"
     CDS             169. .1404
                     /note="ORF1"
                     /codon_start=1
                     /transl_table=11
                     /product="putative integral membrane protein"
                     /protein_id="AAG34708.1"
                     /db_xref="GI:11345437"
                     /translation="MFKPISLYIGLRYTRARRSNHFISFIALVSMIGLTLGVAVLITV
                     LSVMNGFDRELKNRVLGMVPQATISSTQILTNWPELAKKVEGHKHVVGVAPFTQLQGM
                     LTAQGQVAGIMVSGIDPQYEKKVSIIQNHMVAGSIDNLKKGEFGIVLGKQMTDALGLG
                     LNDNVTLVLPEATPSPAGVVPRFKRFKIVGIFSIGAEVDSMMGYIALNDASTLLRLPD
                     GAQGIRLKLDDIFAAPQVADDIVRTLPPNFYASNWTYTHGNLFSAIQMEKAMVSLLLF
                     LIVLVAAFNIVSSLVMVVTDKKSDIAILRTLGASPATITKIFMVQGTVIGVIGTVAGA
                     VLGVIFASSVSNIIDWVNTALGLHLFDAYFINYLPSYLRWQDVVLIVSVSLLLSFLAT
                     IYPALRAAKIQPAEALRYE"
CDS              1397. .2083
                 /note="ORF2"
                 /codon_start=1
                 /transl_table=11
                 /product="putative ATP binding protein of ABC transporter"
                 /protein_id="AAG34709.1"
                 /db_xref="GI:11345438"
                 /translation="MSKIVLEAKDVYKHFTDGKSTVEVIKGLSLKIAAGEFVSIVGAS
                 GSGKSTLLHILGGLDQPTQGQVFLNEQRFDNLGEAERGFKRNQYLGFVYQFHHLLPEF
                 SALENVAMPLMLRADTNYKEVKQQAEHLLDRVGLSHRLTHKPGELSGGERQRVALARA
                 LVARPAVMLADEPTGNLDRKTAFGIFELLSDLKQEFNMAMLIVTHDEQLAQSADSILH
                 MQDGLWVDHS"
gene             2210. .4588
                 /gene="comA"
CDS              2210. .4588
                 /gene="comA"
                 /codon_start=1
                 /transl_table=11
                 /product="competence factor ComA"
                 /protein_id="AAG34710.1"
                 /db_xref="GI:11345439"
                 /translation="MMGRTLECIEPFTCITGILTFLIWAIYFYKKRKHQIGSGLRALV
                 ILISASFLFGIGAQYADSALEERLYYKIMHSDTHEVIIYVHQMGESSDKGIKQVAEVL
                 NIHPDVVNWVFYIDQSLLGAEKSSSLELGQYYRITAKTKPIHSYALEGAFDVEKWFVQ
                 QNWQATAQVYRIQTLSKQEVMSAGFYLHVKQQQTWRAQFLQSVEKLRLKFRNYFIDSS
                 LSHHALLLALLTGDRSLLNPELEQQFQRFGISHLLAISGPHVLIFAAMVSWIFTRLIH
                 RFKPEIYLSIPRQKLVLLPFLMCVILYCAFVGFEIPALRTLFTVMLGSILLLIRHHLS
                 AFSLLVLSASLLLWMDPFSILSAAFWLSYGACFILIRIYQDMTQAQQFHQTRFTRGMQ
                 ILKALVRSQWKIFVALMPLTLIIFKQISWIAPLSNLIAIPILGGIIVPLDIMAGCLWL
                 FIPSLGNLLFYLNNSLLELFIGILKGLDHLFNPQLQYYAFTPLLLICLTLAIVILFLP
                 KGVVPKFWSVLCFVPLLIPSKMIRPTQISILDVGQGQSIFIQDQTQRILLDTGGYYDE
                 SKFSIGKQVVVPFLKSQGVHQLDRVILTHLDQDHSGGLPAVLQLIPTRMIQANEEIKD
                 LSIPFSLCHSGQTWETSELKIQILSPKLENLQSTVFNRNEASCVMYLQLKQASDLKYF
                 LMMGDAGWETEYQLLQQYPDLKVDVLILGHHGSQHSSSYDFLKRYHPKIAIASAGFEN
                 GYGHPSAEVSNRLQDLNIPLLSTTDKGAIHFKFDSGNEQTEFYRQSYKWLQR"
CDS              complement(4521. .5534)
                 /note="ORF4"
                 /codon_start=1
                 /transl_table=11
                 /product="putative HtrB protein"
                 /protein_id="AAG34711.1"
                 /db_xref="GI:11345440"
                 /translation="MTTSDSRSFMYRLLKFISRQPIQLSRLFARMLAGLVNTLKLTKT
                 SRTIQLNLKITLPYLNDQQRERIIQNAIRNELTSYMEFFSIWGSSNQKNISRIHHVEG
                 ADFFHQAIAENKGIVLIVPHFGTWEIMNAWCAQYTDMTILYKPVKDKDADRFVREARS
                 REQAKLVPTDESGVRQIFKALKQGGTTVILPDHTPNVGGEMVDYFGIPLASSNLSAKL
                 IQKTKAKALFLYAIRNENHGFDMFIEPMDPAIYQGTDHDGTLVIHHAIEDLIRKYPDH
                 YHWSYKRFKANPELDNLYTLNEEDALANVEQVRQANRERLNAATTYTTDDKILSAHSL
                 NQT"
CDS              5668. .6684
                 /note="ORF5"
                 /codon_start=1
                 /transl_table=11
                 /product="putative protease"
                 /protein_id="AAG34712.1"
                 /db_xref="GI:11345441"
                 /translation="MSDWPPKPQNEVQNSNNKTGQEWQILEKAVLASVEEQRRTRRWG
                 IFFKFLTFLYLLFVFVLLGRGCTTSESSGGTSSSTAHLAVVNIIGTIDSSNQSVNSED
                 TNKALERAFEAKNSKAVALNINSPGGSPVQSDEIWQEIRYLKKQHPDKKVYAVIGDMG
                 ASGAYYIASAADEIIVNPSSLVGSIGVIMPNYGITGLAQKLGIEDRTLTSGSNKDILS
                 MTKPVNPAQREHIQSVLDNVHAHFINAVKEGRGTRLKSNDPAIFSGLFWTGEQAIQLG
                 VADRTGNLDTLMRELKVENKVNYTIERSPFESILGKMGSELGQGISQSVSQQLKTEQQ
```

```
                     AKLQ"
     CDS             6696. .7502
                     /note="ORF6"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAG34713.1"
                     /db_xref="GI:11345442"
                     /translation="MQNKPIIHFAHANGVPSKVYQKLFDQLSDQYEVIYVPLIGPDKR
                     YPVTDHWPHLVDQIIDSIVRQAGGRQVIGLGHSLGSVLTMMASYKRPELFSQVIMLDP
                     PLILGIHSLGLHLAKHLSPKTVDRMTPAGLSAHRRDHWESRQQAYELLRSKGFYKNFD
                     EDCFQAYIDYALTDDAVRGGVTLTIPKNDEVMLFRTNPSWWWLTLHRACSVPIDLVVG
                     AQSDFLKRGIPQKAKKKLGIPFTVVGRWAYVSPLNIRRETVNQIQKLIRA"
BASE COUNT      2151 a    1377 c    1688 g    2482 t
ORIGIN


alignment_scores:
            Quality:    7.00              Length:      7
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_236_243 x AF320001    ..

 Align seg 1/1  to: AF320001  from: 1  to: 7698

       2 AspSerGluValAspSerSer 8
         ||||||||||||||||||||
    7478 GATTCAGAAGTTGATTCGAGC 7498

seq_name: gb_ba3:MBDR3S1

seq_documentation_block:
LOCUS        MBDR3S1      9281 bp    DNA              BCT       27-APR-1996
DEFINITION   Mycobacterium bovis deletion region 3, 5' end.
ACCESSION    U35017
VERSION      U35017.1  GI:1049238
KEYWORDS     .
SEGMENT      1 of 2
SOURCE       Mycobacterium bovis.
  ORGANISM   Mycobacterium bovis
             Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
             Actinomycetales; Corynebacterineae; Mycobacteriaceae;
             Mycobacterium; Mycobacterium tuberculosis complex.
REFERENCE    1  (bases 1 to 9281)
  AUTHORS    Mahairas,G.G., Sabo,P.J., Hickey,M.J., Singh,D.C. and Stover,C.K.
  TITLE      Molecular analysis of genetic differences between Mycobacterium
             bovis BCG and virulent M. bovis
  JOURNAL    J. Bacteriol. 178 (5), 1274-1282 (1996)
  MEDLINE    96200095
REFERENCE    2  (bases 1 to 9281)
  AUTHORS    Mahairas,G.G., Sabo,P.J., Hickey,M.J., Singh,D.C. and Stover,C.K.
  TITLE      Direct Submission
  JOURNAL    Submitted (29-AUG-1995) Mark J. Hickey, Molecular Microbiology,
             PathoGenesis Corp., 201 Elliott Ave. W., Seattle, WA 98119, USA
FEATURES             Location/Qualifiers
     source          1. .9281
                     /organism="Mycobacterium bovis"
                     /db_xref="taxon:1765"
     misc_difference replace(1400. .>9281,"")
                     /note="absent in related avirulent strain M. bovis BCG"
BASE COUNT      1706 a    3191 c    2929 g    1453 t       2 others
ORIGIN


alignment_scores:
            Quality:    7.00              Length:      7
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_236_243 x MBDR3S1/rev    ..

 Align seg 1/1  to reverse of: MBDR3S1  from: 1  to: 9281

       2 AspSerGluValAspSerSer 8
         ||||||||||||||||||||
    1452 GATTCGGAGGTAGACAGCAGC 1432

seq_name: gb_pat2:I86264

seq_documentation_block:
LOCUS        I86264       12412 bp    DNA              PAT       10-JUN-1998
DEFINITION   Sequence 18 from patent US 5700683.
ACCESSION    I86264
VERSION      I86264.1  GI:3205982
KEYWORDS     .
SOURCE       Unknown.
  ORGANISM   Unknown.
             Unclassified.
REFERENCE    1  (bases 1 to 12412)
  AUTHORS    Stover,C.Kendall and Mahairas,G.G.
  TITLE      Virulence-attenuating genetic deletions deleted from mycobacterium
             BCG
  JOURNAL    Patent: US 5700683-A 18 23-DEC-1997;
FEATURES             Location/Qualifiers
     source          1. .12412
                     /organism="unknown"
BASE COUNT      2146 a    4274 c    3875 g    1917 t     200 others
ORIGIN


alignment_scores:
            Quality:    7.00              Length:      7
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_236_243 x I86264/rev    ..

 Align seg 1/1  to reverse of: I86264  from: 1  to: 12412

       2 AspSerGluValAspSerSer 8
         ||||||||||||||||||||
    1459 GATTCGGAGGTAGACAGCAGC 1439

seq_name: gb_ba3:MTCY336

seq_documentation_block:
LOCUS        MTCY336      32437 bp    DNA              BCT       24-JUN-1999
DEFINITION   Mycobacterium tuberculosis H37Rv complete genome; segment 70/162.
ACCESSION    Z95586 AL123456
VERSION      Z95586.1  GI:3261785
KEYWORDS     .
SOURCE       Mycobacterium tuberculosis.
  ORGANISM   Mycobacterium tuberculosis
             Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
             Actinomycetales; Corynebacterineae; Mycobacteriaceae;
             Mycobacterium; Mycobacterium tuberculosis complex.
REFERENCE    1  (bases 1 to 32437)
  AUTHORS    Cole,S.T., Brosch,R., Parkhill,J., Garnier,T., Churcher,C.,
             Harris,D., Gordon,S.V., Eiglmeier,K., Gas,S., Barry III,C.E.,
             Tekaia,F., Badcock,K., Basham,D., Brown,D., Chillingworth,T.,
             Connor,R., Davies,R., Devlin,K., Feltwell,T., Gentles,S.,
             Hamlin,N., Holroyd,S., Hornsby,T., Jagels,K., Krogh,A., McLean,J.,
             Moule,S., Murphy,L., Oliver,S., Osborne,J., Quail,M.A.,
             Rajandream,M.A., Rogers,J., Rutter,S., Seeger,K., Skelton,S.,
             Squares,S., Sqares,R., Sulston,J.E., Taylor,K., Whitehead,S. and
             Barrell,B.G.
  TITLE      Deciphering the biology of Mycobacterium tuberculosis from the
             complete genome sequence
  JOURNAL    Nature 393 (6685), 537-544 (1998)
  MEDLINE    98295987
  REMARK     Erratum:[[published erratum appears in Nature 1998 Nov
             12;396(6707):190]]
```

```
REFERENCE   2  (bases 1 to 32437)
  AUTHORS   Parkhill,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium
            tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome
            Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unite de Genetique
            Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux,
            75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk
COMMENT     On Jun 27, 1998 this sequence version replaced gi:2117233.
            Notes:
            Details of M. tuberculosis sequencing at the Sanger Centre are
            available on the World Wide Web.
            (URL, http://www.sanger.ac.uk/Projects/M_tuberculosis/) CDS have
            been renumbered from the original cosmid submissions but the old
            gene designations are in brackets after the new gene numbers.
            Gene prediction was based on a Hidden Markov Model of TB genes
            implemented in TBparse (Krogh) supplemented with visual inspection
            of positional base preference in codons, especially where there is
            an increase in the observed/expected third position G + C.
            CAUTION:  In some cases we may not have predicted the correct
            initiation codon.  Where possible we choose an initiation codon
            (atg, gtg, or ttg) which is preceded by an upstream ribosome
            binding site sequence (optimally 5-13bp before the initiation
            codon).  If this cannot be identified we choose the most upstream
            initiation codon.
FEATURES             Location/Qualifiers
     source          1. .32437
                     /organism="Mycobacterium tuberculosis"
                     /strain="H37Rv"
                     /db_xref="taxon:1773"
     source          <1. .273
                     /organism="Mycobacterium tuberculosis"
                     /strain="H37Rv"
                     /db_xref="taxon:1773"
                     /clone="Y48"
     source          33. .32091
                     /organism="Mycobacterium tuberculosis"
                     /strain="H37Rv"
                     /db_xref="taxon:1773"
                     /clone="Y336"
     gene            complement(39. .2228)
                     /gene="Rv1565c"
     CDS             complement(39. .2228)
                     /gene="Rv1565c"
                     /note="Rv1565c, (MTCY336.38), len: 729. Unknown membrane
                     protein, some similarity to O05402 HYPOTHETICAL 72.2 KD
                     PROTEIN from B. subtilis (634 aa) opt: 384 E(): 4.8e-17;
                     29.1% identity in 378 aa overlap and to hypothetical
                     protein in H.influenzae. N-terminal half hydrophobic.
                     FASTA results, Y392_HAEIN P43993 hi0392 (245 aa) opt: 265;
                     E(): 5.5e-10; 28.3% identity in 247 aa overlap. TBparse
                     score is 0.930"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein Rv1565c"
                     /protein_id="CAB09072.1"
                     /db_xref="GI:2117272"
                     /db_xref="SPTREMBL:O06625"
                     /translation="MLTLSPPRPPALTPEPALPPVTMGTRTTGFYRHDLDGLRGVAIA
                     LVAVFHVWFGRVSGGVDVFLALSGFFFGGKILRAALNPDLSLSPIAEVIRLIRRLLPA
                     LVVVLAGCALLTIAIQPQTRWEAFANQSLASLGYYQNWELASTVSNYLRAGEAVSPLQ
                     HIWSMSVQGQFYLAFLLLVAGCAYLLRRLFRGPRAPYLRTMFVVLLSTLTLASFIYAI
                     VAHHAYQATAYYNTFARAWELLAGALVGAVVPHVRWPMWLRTAVATAALAAILSCGAL
                     IDGVKEFPGPWALVPVGATMLMILAGANRQGHPGTRDRLPLPNRLLATAPLVALGAMA
                     YSWYLWHWPLLIFWLSYTGHRHANFVEGAAVLLVSGLLAYLTTRLVEDPLRYRAPAGV
                     RSPAAVPPIPWRLRLRRPTIVLGSVVALLGVALTATSFTWREHVIVQRAAGKELSGLS
                     SRDYPGARALIDHVRVPKLRMRPTVLEVRHDLPTSTKDGCISDFVNPAIINCTYGDVD
                     APRTIALAGGSHAEHWLTALDLLGRMHHFKVVTYLKMGCPLSTEEVPLIMGNNAPYPQ
                     CHQWVQAAMAKLVADHPDYVFTTSTRPWNIKPGDVMPATYVGIWQTFADNNIPVLAMR
                     DTPWLVKDGQPFIPADCLAKGGNPQSCGIARSKVLVDRNPTLDFVARFPLLKPLDMSD
                     AICRTDTCRAVEGNVLVYRDSHHLTPTYMRTMTSELGRQIAANTDWW"
     gene            complement(2327. .3019)
                     /gene="Rv1566c"
     CDS             complement(2327. .3019)
                     /gene="Rv1566c"
                     /note="Rv1566c, (MTCY336.37), len: 230. Function: unknown
                     probably exported has QQAPV repeats at C-terminus, similar
                     to hypothetical M. leprae protein, to putative invasins
                     1,2(O07390, O07391) from M. avium and slightly similar to
                     C-terminus of Listeria invasion-associated protein p60
                     precursor P60_LISMO P21171. FASTA results, Q49634 COSMID
                     B1170 (246aa) opt: 957 E(): 0; 70.0% identity in 207 aa
                     overlap. TBparse scoreis 0.890"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein Rv1566c"
                     /protein_id="CAB09071.1"
                     /db_xref="GI:2117271"
                     /db_xref="SPTREMBL:O06624"
                     /translation="MKRSMKSGSFAIGLAMMLAPMVAAPGLAAADPATRPVDYQQITD
                     VVIARGLSQRGVPFSWAGGGISGPTRGTGTGINTVGFDASGLIQYAYAGAGLKLPRSS
                     GQMYKVGQKVLPQQARKGDLIFYGPEGTQSVALYLGKGQMLEVGDVVQVSPVRTNGMT
                     PYLVRVLGTQPTPVQQAPVQPAPVQQAPVQQAPVQQAPVQQAPVQQAPVQQAPVQQAP
                     VQPPPFGTARSR"
     RBS             complement(3030. .3036)
                     /note="possible RBS, AAGGAGG, for Rv1566c"
     misc_feature    complement(3258. .3263)
                     /note="ASNI site: ATTAAT; probably linking fragments
                     B3/G6"
     gene            complement(3259. .3543)
                     /gene="Rv1567c"
     CDS             complement(3259. .3543)
                     /gene="Rv1567c"
                     /note="Rv1567c, (MTCY336.36), len: 94. Unknown. TBparse
                     score is 0.915"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein Rv1567c"
                     /protein_id="CAB09070.1"
                     /db_xref="GI:2117270"
                     /db_xref="SPTREMBL:O06623"
                     /translation="MVTMTSWPSRLFAFTDNVCPPDACPLVPFGVNYYIYPVMWGGIG
                     AAIATAVIGPFVSMLKGWYMSFWPIISIAVITVTSIAGYAIAGFSERYWH"
     gene            3791. .5104
                     /gene="bioA"
     CDS             3791. .5104
                     /gene="bioA"
                     /note="Rv1568, (MTCY336.35c), bioA, len: 437. Function:
                     probable adenosylmethionine-8-amino-7-oxononanoate
                     aminotransferase (EC2.6.1.62) O06622. Contains
                     aminotransferases class-III pyridoxal-phosphate attachment
                     site (PS00600). FASTAresults, BIOA_MYCLE P4548 (436 aa)
                     opt: 2534; E(): 0; 85.1%identityin 436 aa overlap. Also
                     similar to other M. tuberculosis proteins e.g.
                     MTCY227.12c, (449 aa). FASTA score: E():3.5e-16; 29.5%
                     identity in 421 aa overlap. TBparse scoreis0.874"
                     /codon_start=1
                     /transl_table=11
                     /product="bioA"
                     /protein_id="CAB09069.1"
                     /db_xref="GI:2117269"
                     /db_xref="SWISS-PROT:O06622"
                     /translation="MAAATGGLTPEQIIAVDGAHLWHPYSSIGREAVSPVVAVAAHGA
                     WLTLIRDGQPIEVLDAMSSWWTAIHGHGHPALDQALTTQLRVMNHVMFGGLTHEPAAR
                     LAKLLVDITPAGLDTVFFSDSGSVSVEVAAKMALQYWRGRGLPGKRRLMTWRGGYHGD
                     TFLAMSICDPHGGMHSLWTDVLAAQVFAPQVPRDYDPAYSAAFEAQLAQHAGELAAVV
                     VEPVVQGAGGMRFHDPRYLHDLRDICRRYEVLLIFDEIATGFGRTGALFAADHAGVSP
                     DIMCVGKALTGGYLSLAATLCTADVAHTISAGAAGALMHGPTFMANPLACAVSVASVE
                     LLLGQDWRTRITELAAGLTAGLDTARALPAVTDVRVCGAIGVIECDRPVDLAVATPAA
                     LDRGVWLRPFRNLVYAMPPYICTPAEITQITSAMVEVARLVGSLP"
     misc_feature    4541. .4654
                     /gene="bioA"
                     /note="PS00600 Aminotransferases class-III
                     pyridoxal-phosphate attachment site"
     gene            5101. .6261
                     /gene="bioF"
```

```
     CDS             5101. .6261
                     /gene="bioF"
                     /note="Rv1569, (MTCY336.34c), bioF, len: 386. Function:
                     probable 8-amino-7-oxononanoate synthase (EC 2.3.1.47)
                     O06621. Contains aminotransferases class-II
                     pyridoxal-phosphate attachment site (PS00599). FASTA
                     results, BIOF_MYCLE P45487(385 aa) opt: 1971; E(): 0;
                     80.1% identity in 381 aa overlap. Also similar to
                     MTCY10H4.32, FASTA score: E(): 5.5e-29;37.4% identity in
                     393 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="bioF"
                     /protein_id="CAB09068.1"
                     /db_xref="GI:2117268"
                     /db_xref="SWISS-PROT:O06621"
                     /translation="MKAATQARIDDSPLAWLDAVQRQRHEAGLRRCLRPRPAVATELD
                     LASNDYLGLSRHPAVIDGGVQALRIWGAGATGSRLVTGDTKLHQQFEAELAEFVGAAA
                     GLLFSSGYTANLGAVVGLSGPGSLLVSDARSHASLVDACRLSRARVVVTPHRDVDAVD
                     AALRSRDEQRAVVVTDSVFSADGSLAPVRELLEVCRRHGALLLVDEAHGLGVRGGGRG
                     LLYELGLAGAPDVVMTTTLSKALGSQGGVVLGPTPVRAHLIDAARPFIFDTGLAPAAV
                     GAARAALRVLQAEPWRPQAVLNHAGELARMCGVAAVPDSAMVSVILGEPESAVAAAAA
                     CLDAGVKVGCFRPPTVPAGTSRLRLTARASLNAGELELARRVLTDVLAVARR"
     misc_feature    5806. .5835
                     /gene="bioF"
                     /note="PS00599 Aminotransferases class-II
                     pyridoxal-phosphate attachment site"
     gene            6258. .6938

alignment_scores:
            Quality:    7.00              Length:      7
              Ratio:  1.000               Gaps:        0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_236_243 x MTCY336   ..

 Align seg 1/1  to: MTCY336  from: 1  to: 32437

       2 AspSerGluValAspSerSer 8
         ||||||||||||||||||||
   16860 GATTCGGAGGTAGACAGCAGC 16880

seq_name: gb_pl1:AB026650

seq_documentation_block:
LOCUS        AB026650    64569 bp    DNA             PLN       27-DEC-2000
DEFINITION   Arabidopsis thaliana genomic DNA, chromosome 5, P1 clone:MPF21.
ACCESSION    AB026650 BA000015
VERSION      AB026650.1  GI:4757406
KEYWORDS     .
SOURCE       Arabidopsis thaliana (strain:Columbia) DNA, clone_lib:Mitsui P1
             clone:MPF21.
  ORGANISM   Arabidopsis thaliana
             Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
             Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
             Brassicales; Brassicaceae; Arabidopsis.
REFERENCE    1  (sites)
  AUTHORS    Kaneko,T., Katoh,T., Asamizu,E., Sato,S., Nakamura,Y., Kotani,H.
             and Tabata,S.
  TITLE      Structural analysis of Arabidopsis thaliana chromosome 5. XI
  JOURNAL    Unpublished
REFERENCE    2  (bases 1 to 64569)
  AUTHORS    Nakamura,Y.
  TITLE      Direct Submission
  JOURNAL    Submitted (28-APR-1999) Yasukazu Nakamura, Kazusa DNA Research
             Institute, Department of Plant Gene Research; 1532-3, Yana,
             Kisarazu, Chiba 292-0812, Japan (E-mail:ynakamu@kazusa.or.jp,
             Tel:81-438-52-3935, Fax:81-438-52-3934)
COMMENT      Address for correspondence: kaos@kazusa.or.jp
             For the latest information on annotation of this clone, please see
             http://www.kazusa.or.jp/kaos/cgi-bin/agd_graph.cgi?c=MPF21
             Genes with similarity to proteins in the databases are described in
             'product' or 'note' qualifiers.  Genes that have no significant
             protein similarity are described as 'unknown protein'.
             The software programs used to predict genes include: Grail
             (Informatics Group, Oak Ridge National Laboratory,
             http://compbio.ornl.gov/Grail-1.3/),
             GENSCAN (Chris Burge, MIT, http://CCR-081.mit.edu/GENSCAN.html),
             NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of
             Denmark, http://www.cbs.dtu.dk/services/NetGene2/) and
             SplicePredictor (Volker Brendel, Stanford University,
             http://gremlin1.zool.iastate.edu/cgi-bin/sp.cgi).
             Genes encoding tRNAs are predicted by tRNAscan-SE
             (Sean Eddy, Washington University School of Medicine, St. Louis,
             http://genome.wustl.edu/eddy/tRNAscan-SE/).
             This sequence may not be the entire insert of this clone. It may be
             shorter because we remove overlaps between neighboring submissions.
             The 5' clone is K9P8 and the 3' clone is K6A12.
FEATURES             Location/Qualifiers
     source          1. .64569
                     /organism="Arabidopsis thaliana"
                     /strain="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="5"
                     /clone="MPF21"
                     /clone_lib="Mitsui P1"
     CDS             complement(join(2128. .2334,2532. .2666,2740. .2914,
                     2994. .3190,3282. .3403,4002. .4323))
                     /note="gene_id:MPF21.1"
                     /codon_start=1
                     /evidence=not_experimental
                     /product="protein kinase"
                     /protein_id="BAB10286.1"
                     /db_xref="GI:10177211"
                     /translation="MKEGKDGFVRADQIDLKSLDEQLERHLSRALTLEKNKKKDEEDT
                     TAVAIGGSASSSPVTLNGGGFVGKRKQRLEWEIDPSKLIIKTVLARGTFGTVHRGIYD
                     GQDVAVKLLDWGEEGHRSEAEIVSLRADFAQEVAVWHKLDHPNVTKFIGATMGASGLQ
                     LQTESGPLAMPNNICCVVVEYLPGGALKSYLIKNRRRKLTFKIVVQLALDLARGLSYL
                     HSQKIVHRDVKTENMLLDKTRTVKIADFGVARVEASNPNDMTGETGTLGYMAPEVLNG
                     NPYNRKCDVYSFGICLWEIYCCDMPYPDLTFSEVTSAVVRQNLRPDIPRCCPSALAAV
                     MKRCWDANPDKRPEMDEVVPMLESIDTTKGGGMIPNDQQQGCLCFRRKRGP"
     CDS             8846. .9781
                     /note="gene_id:MPF21.2
                     unknown protein"
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="BAB10287.1"
                     /db_xref="GI:10177212"
                     /translation="MGQDRGFGFPTQRLCSLSSLALSHLGKQDLNLVSKTCGDTTDMF
                     STRGSYQVSTQVSQSYFDGYCGWVHGSSHLQQQFLPPQNQCMKQVPLQVDGVISKAEE
                     QCSQKRFLVFDQSGDQTTLLLASDIRKSFETLKQHACPDMKEELQRSNKDLFVCHGMQ
                     GNSEPDLKEDSEELNALLYSEDESGYCSEEDEVTSADHSPSIVVSGREDQKTFLGSYG
                     QPLNAKKRKILETSNESMRDAESSCGSCDNTRISFLKRSKLSSNKIGEEKIFETVSLL
                     RSVVPGEELVDPILVIDRAIDYLKSLKMEAKNREA"
     CDS             join(10881. .10919,11556. .11741,12132. .12431,12620. .12857,
                     13029. .13414)
                     /note="contains similarity to unknown protein
                     gene_id:MPF21.3
                     pir||T05571"
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="BAB10288.1"
                     /db_xref="GI:10177213"
                     /translation="MAGRVFEAWKGSNKFLFGGRLIFGPDAWSIPFTFLLIITPVCFF
                     SVFVATHLRRELLPNNAGHVFLVAGVLFTVFVLILLFLTSARDPGIVPRNSHPPEEEL
                     CYDTTVSSDGRQTPTVQIPRTKEVMVYGVSVRVKYCDTCMLYRPPRCSHCSICNNCVE
                     RFDHHCPWVGQCIGVRNYRYFFMFVSSATILCIYIFSMSALYIKVLMDNHQGTVWRAM
                     RESPWAVMLMIYCFISLWFVGGLTGFHLYLISTNQVKPSRNDFRAFIKEEPPRNITLA
                     TTWERPEEADEENREERRQKVEDDLDIDEDVMKLQQRLNDEEGSDTAHHKIDIDQMRI
                     GSNERAPTIRSEARHGNWGARSNAQEEDVIAGSSVRESRSYAAAEEGR"
     CDS             join(13956. .14141,14362. .14739)
                     /note="contains similarity to pollen-specific protein
                     Bnm1
                     gene_id:MPF21.4"
                     /codon_start=1
```

```
                     /evidence=not_experimental
                     /protein_id="BAB10289.1"
                     /db_xref="GI:10177214"
                     /translation="MASYSGCFFVVTLAVLSQFLLAPASALVNREYIDSNCQRVKNKA
                     FCIQTLTTYPPAAAATGLLPLAEAVVGLAISHCEKTAGFAAETAKKDATLKTQFNECH
                     DAYVGILASLKSALLELKDSPDTANYDVMVSGDDTRRVKGLVEKNTDTASKTLMEMTL
                     QMEQLLDLAAGATDAADDDDENLHRRA"
CDS                  join(16416. .16592,17047. .17397)
                     /note="contains similarity to pollen-specific protein
                     Bnm1
                     gene_id:MPF21.5"
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="BAB10290.1"
                     /db_xref="GI:10177215"
                     /translation="MASPSNYFFLVSLAVLLPFLIIFASATTYIDAICQSVTDKAFCA
                     KTLNAYPAAGSATSQFQAATATLNLAISYADKSAGFTGNAAKENPTLKTQFAASQDAF
                     VTISKSLKSAASELKISPDTANYDVMVCSDSIATVKNLVEKNTDNSSKTVMTMTLMME
                     KLLTIAVGATVAVGG"
CDS                  join(19322. .19498,19828. .20178)
                     /note="gene_id:MPF21.6
                     pir||T05570
                     similar to unknown protein"
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="BAB10291.1"
                     /db_xref="GI:10177216"
                     /translation="MASSSYGFGVVYLAVLLPCLLVPASATKYIDAICHLVSDKAFCT
                     KTLNAYPPATSATSTFEAAVATLHLGVSYANNAAGFAGEAAKENPQLKEKFDACQLEL
                     ALISRGLKNASEELYDDTAYFDVTDCLDNIATMKNLVGKNTDKASKTIITMTLMLENF
                     IFLAIGATSTIDIKT"
CDS                  join(21840. .21989,22401. .22751)
                     /note="contains similarity to pollen-specific protein
                     Bnm1
                     gene_id:MPF21.7"
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="BAB10292.1"
                     /db_xref="GI:10177217"
                     /translation="MAVLSEILFISSSGLTSTKYIDALCQHVENGTFCMHMLRAYPPA
                     VTATDLFSLANIVINLGISYANNTGAFAAEAAKKEPTLKVQFKSFQHEYDGIQLYLRM
                     ARGELKESPMSANYDIFRCTDITMYLKDVVGGNRDNTSKTFMDMTLQMENIIDVALGA
                     TVALGG"
CDS                  join(23846. .24043,24366. .24719)
                     /note="contains similarity to pollen-specific protein
                     Bnm1
                     gene_id:MPF21.8"
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="BAB10293.1"
                     /db_xref="GI:10177218"
                     /translation="MTSSSHCFLIVSVALLLQVLLSPVSALTLTSTKYINKLCQMSFV
                     DNKPFCLQTLSAYPPAASATGLLPLASAVIRGIVLPQAKKSASFANGAAKKEPSLKTQ
                     FQTCQEAFVSIGDSLMSAFEELKRDPHIATYDVMTCTDHTTQVKNLVGKNTDVASKTL
                     IEMTMQMEKLIPLALASTQVLGG"
CDS                  join(25238. .25439,25665. .26125)
                     /note="contains similarity to AP2 domain transcription
                     factor
                     gene_id:MPF21.9"
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="BAB10294.1"
                     /db_xref="GI:10177219"
                     /translation="MSAMVSALTQVVSARSQTEAEGAHSQTEAEGAHSSSSSAGHKRG
                     WLGIDSAPIPSSFARVDSSHNPIEEAREKKRRYRGVRQRPWGKWAAEIRDPHRAARVW
                     LGTFDTAEAAARAYDEAALRFRGNKAKLNFPEDVRILPPPPPLLRSPADTVANKAEED
                     LINYWSYTKLLQSSGQRSFLERGQEESSNIFEHSPMEQPLPPSSSGPSSSNFPAPSLP
                     NT"
CDS                  join(29251. .29515,29877. .29998,30076. .30168)
                     /note="gene_id:MPF21.10
                     unknown protein"
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="BAB10295.1"
                     /db_xref="GI:10177220"
                     /translation="MGNCQAVDTARVVIQHPNGKEEKLSCPVSASYVMKMNPGHCVSL
                     LISTTALSSASSGHGGPLRLTRIKLLRPTDTLVLGHVYRLITTKEVMKGLMAKKCSKL
                     KKESKGSDDKLEMVKAINSTKLDNEDQLQMKKQEKERSRISRSWQPSLQSISEGGSS"
CDS                  join(31553. .31591,31679. .31786,31876. .31932,32002. .32067,
                     32152. .32212,32296. .32393)
                     /note="gene_id:MPF21.11
                     pir||S75732
                     similar to unknown protein"
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="BAB10296.1"
                     /db_xref="GI:10177221"
                     /translation="MLYDGDCPLCMREVNMLMERNEKHGTIKFVDISSNDYSPEDNQG
                     LDYKTVMGQIHAIQSDGNVVKGVEAFRRLYEEVGLGWVYTITKFEPIGKLADVVYDVW
                     AKYRLQVTGRPSIEAILEARKKDKVGNKGYRFSVVVTFAR"
CDS                  join(32805. .33053,33489. .33681,33777. .33886,34354. .34432,
                     34531. .34598,34773. .34839,35246. .35332,35645. .35781)
                     /note="contains similarity to glucose inhibited division
                     protein B
                     gene_id:MPF21.12"
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="BAB10297.1"
                     /db_xref="GI:10177222"
                     /translation="MRSICSFVENTKKMHCVSHILSLGHLIKHLPCTKSHTFLRRRFH

alignment_scores:
            Quality:    7.00              Length:      7
              Ratio:   1.000               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_236_243 x AB026650/rev   ..

 Align seg 1/1  to reverse of: AB026650  from: 1  to: 64569

       2 AspSerGluValAspSerSer 8
         ||||||||||||||||||||
   44873 GACAGTGAAGTTGACTCATCA 44853

seq_name: gb_pr2:AC008530

seq_documentation_block:
LOCUS        AC008530    89095 bp    DNA             PRI       31-MAY-2000
DEFINITION   Homo sapiens chromosome 19 clone CTC-478M6, complete sequence.
ACCESSION    AC008530
VERSION      AC008530.7  GI:8122100
KEYWORDS     HTG.
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 89095)
  AUTHORS    DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE      Direct Submission
  JOURNAL    Unpublished
REFERENCE    2  (bases 1 to 89095)
  AUTHORS    DOE Joint Genome Institute.
  TITLE      Direct Submission
  JOURNAL    Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
             Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE    3  (bases 1 to 89095)
  AUTHORS    DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE      Direct Submission
  JOURNAL    Submitted (31-MAY-2000) DOE Joint Genome Institute, 2800 Mitchell
             Drive, Walnut Creek, CA 94598, USA
COMMENT      On May 31, 2000 this sequence version replaced gi:7711288.
             Draft Sequence Produced by DOE Joint Genome Institute
             www.jgi.doe.gov
             Finishing Completed at Stanford Human Genome Center
             www-shgc.stanford.edu
```

```
            Quality: Phrap Quality >=40 99.9% of Sequence;
            Estimated Total Number of Errors is 0.1.
FEATURES                Location/Qualifiers
     source             1. .89095
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /chromosome="19"
                        /clone="CTC-478M6"
BASE COUNT     27679 a  18534 c  17203 g  25679 t
ORIGIN


alignment_scores:
             Quality:    7.00              Length:      7
               Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_236_243 x AC008530   ..

 Align seg 1/1  to: AC008530  from: 1  to: 89095

       1 ThrAspSerGluValAspSer 7
         |||||||||||||||||||||
   84199 ACAGACTCTGAGGTCGACTCA 84219

seq_name: gb_htg24:AP002018

seq_documentation_block:
LOCUS       AP002018   125674 bp    DNA              HTG       01-JUN-2000
DEFINITION  Homo sapiens chromosome 4 clone 2058G14 map 4q22-q24, ***
            SEQUENCING IN PROGRESS ***, 19 unordered pieces.
ACCESSION   AP002018
VERSION     AP002018.1  GI:7798578
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Homo sapiens DNA, clone:2058G14.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 125674)
  AUTHORS   Tsai,S.F.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-MAY-2000) Shih-Feng Tsai, National Yang-Ming
            University, Institute of Genetics; 155 Li-Rong St. Section 2,
            Peitou, Taipei, Taiwan 11221, Republic of China
            (E-mail:ympetsai@ym.edu.tw, URL:http://genome.ym.edu.tw/,
            Tel:886-2-28267043, Fax:886-2-28264930)
COMMENT     Gaps between the contigs are represented as 100 N.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 19 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1775: contig of 1775 bp in length
            *     1776 1875: gap of      100 bp
            *     1876     3417: contig of 1542 bp in length
            *     3418 3517: gap of      100 bp
            *     3518     4578: contig of 1061 bp in length
            *     4579 4678: gap of      100 bp
            *     4679     6139: contig of 1461 bp in length
            *     6140 6239: gap of      100 bp
            *     6240     7815: contig of 1576 bp in length
            *     7816 7915: gap of      100 bp
            *     7916    11064: contig of 3149 bp in length
            *    11065 11164: gap of      100 bp
            *    11165    14678: contig of 3514 bp in length
            *    14679 14778: gap of      100 bp
            *    14779    16592: contig of 1814 bp in length
            *    16593 16692: gap of      100 bp
            *    16693    20750: contig of 4058 bp in length
            *    20751 20850: gap of      100 bp
            *    20851    24893: contig of 4043 bp in length
            *    24894 24993: gap of      100 bp
            *    24994    32048: contig of 7055 bp in length
            *    32049 32148: gap of      100 bp
            *    32149    38202: contig of 6054 bp in length
            *    38203 38302: gap of      100 bp
            *    38303    44488: contig of 6186 bp in length
            *    44489 44588: gap of      100 bp
            *    44589    54348: contig of 9760 bp in length
            *    54349 54448: gap of      100 bp
            *    54449    62197: contig of 7749 bp in length
            *    62198 62297: gap of      100 bp
            *    62298    70876: contig of 8579 bp in length
            *    70877 70976: gap of      100 bp
            *    70977    83901: contig of 12925 bp in length
            *    83902 84001: gap of      100 bp
            *    84002    98894: contig of 14893 bp in length
            *    98895 98994: gap of      100 bp
            *    98995   125674: contig of 26680 bp in length.
FEATURES                Location/Qualifiers
     source             1. .125674
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /chromosome="4"
                        /map="4q22-q24"
                        /clone="2058G14"
BASE COUNT     37414 a  25313 c  23799 g  37308 t   1840 others
ORIGIN


alignment_scores:
             Quality:    7.00              Length:      7
               Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_236_243 x AP002018   ..

 Align seg 1/1  to: AP002018  from: 1  to: 125674

       2 AspSerGluValAspSerSer 8
         ||||||||||||||||||||||
  103729 GACTCTGAAGTTGACTCTTCT 103749

seq_name: gb_htg5:AC016212

seq_documentation_block:
LOCUS       AC016212   160217 bp    DNA              HTG       16-MAR-2000
DEFINITION  Homo sapiens clone RP11-25P7, WORKING DRAFT SEQUENCE, 9 unordered
            pieces.
ACCESSION   AC016212
VERSION     AC016212.3  GI:7249075
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 160217)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone RP11-25P7
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 160217)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
```

```
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-NOV-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 16, 2000 this sequence version replaced gi:6649341.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L4619
                Center clone name: 25_P_7
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 155915 bases at least Q40
                Consensus quality: 158008 bases at least Q30
                Consensus quality: 158845 bases at least Q20
                Insert size: 158000; agarose-fp
                Insert size: 159417; sum-of-contigs
                Quality coverage: 5.1 in Q20 bases; agarose-fp
                Quality coverage: 5.1 in Q20 bases; sum-of-contigs
            --------------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 9 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     1800: contig of 1800 bp in length
            *    1801 1900: gap of      100 bp
            *    1901     4600: contig of 2700 bp in length
            *    4601 4700: gap of      100 bp
            *    4701     8787: contig of 4087 bp in length
            *    8788 8887: gap of      100 bp
            *    8888    16192: contig of 7305 bp in length
            *   16193 16292: gap of      100 bp
            *   16293    29642: contig of 13350 bp in length
            *   29643 29742: gap of      100 bp
            *   29743    53221: contig of 23479 bp in length
            *   53222 53321: gap of      100 bp
            *   53322    75019: contig of 21698 bp in length
            *   75020 75119: gap of      100 bp
            *   75120    99739: contig of 24620 bp in length
            *   99740 99839: gap of      100 bp
            *   99840   160217: contig of 60378 bp in length.
FEATURES            Location/Qualifiers
     source         1. .160217
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /clone="RP11-25P7"
                    /clone_lib="RPCI-11 Human Male BAC"
     misc_feature   1. .1800
                    /note="assembly_fragment"
     misc_feature   1901. .4600
                    /note="assembly_fragment"
     misc_feature   4701. .8787
                    /note="assembly_fragment"
     misc_feature   8888. .16192
                    /note="assembly_fragment
                    clone_end:T7
                    vector_side:right"
     misc_feature   16293. .29642
                    /note="assembly_fragment"
     misc_feature   29743. .53221
                    /note="assembly_fragment"
     misc_feature   53322. .75019
                    /note="assembly_fragment
                    clone_end:SP6
                    vector_side:left"
     misc_feature   75120. .99739
                    /note="assembly_fragment"
     misc_feature   99840. .160217
                    /note="assembly_fragment"
BASE COUNT    45570 a  33216 c  33535 g  47092 t    804 others
ORIGIN


alignment_scores:
            Quality:    7.00              Length:      7
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_236_243 x AC016212/rev   ..

 Align seg 1/1  to reverse of: AC016212  from: 1  to: 160217

       1 ThrAspSerGluValAspSer 7
         |||||||||||||||||||||
  148365 ACAGATTCTGAAGTGGACAGT 148345

seq_name: gb_pr2:AC005747

seq_documentation_block:
LOCUS       AC005747   166701 bp    DNA             PRI       17-OCT-1998
DEFINITION  Homo sapiens chromosome 17, clone hRPK.150_K_15, complete sequence.
ACCESSION   AC005747
VERSION     AC005747.1  GI:3766124
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 166701)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 17, clone hRPK.150_K_15
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 166701)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Boutwell,C.,
            Brown,A., Castle,A., Cerny,J., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., Corliss,D., Depayre,E., Devon,K., Dewar,K.,
            Donelan,L., Ferreira,P., FitzHugh,W., Forrest,C., Funke,R.,
            Gage,D., Gardyna,S., Geraigery,K., Grant,G., Hagos,B., Heaford,A.,
            Herena,L., Horton,L., Howland,J.C., Jacotot,L., Jones,C., Kann,L.,
            Karatas,A., Lehoczky,J., Macdonald,P., Marquis,N., McEwan,P.,
            McGurk,A., McKernan,K., Meldrim,J., Molla,M., Morris,W., Morrow,J.,
            Mychaleckyj,J., Nahf,R., Naylor,J., Niloff,M., O'Connor,T.,
            O'Donnell,P., Pavlin,B., Peterson,K., Riley,R., Roberts,D., Roy,A.,
            Severy,P., Stange-Thomann,N., Stilwell,J., Stojanovic,N., Stone,C.,
            Subramanian,A., Tesfaye,S., Tichovolsky,N., Torruella-Miller,I.,
            Vassiliev,H., Vo,A., Wagner,A., Wheeler,J., Wu,Y., Wyman,D.,
            Ye,W.J., Zhao,J. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-OCT-1998) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 166701)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Boutwell,C.,
            Brown,A., Castle,A., Cerny,J., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., Corliss,D., Depayre,E., Devon,K., Dewar,K.,
            Donelan,L., Ferreira,P., FitzHugh,W., Forrest,C., Funke,R.,
            Gage,D., Gardyna,S., Geraigery,K., Grant,G., Hagos,B., Heaford,A.,
            Herena,L., Horton,L., Howland,J.C., Jacotot,L., Jones,C., Kann,L.,
            Karatas,A., Lehoczky,J., Macdonald,P., Marquis,N., McEwan,P.,
```

```
            McGurk,A., McKernan,K., Meldrim,J., Molla,M., Morris,W., Morrow,J.,
            Mychaleckyj,J., Nahf,R., Naylor,J., Niloff,M., O'Connor,T.,
            O'Donnell,P., Pavlin,B., Peterson,K., Riley,R., Roberts,D., Roy,A.,
            Severy,P., Stange-Thomann,N., Stilwell,J., Stojanovic,N., Stone,C.,
            Subramanian,A., Tesfaye,S., Tichovolsky,N., Torruella-Miller,I.,
            Vassiliev,H., Vo,A., Wagner,A., Wheeler,J., Wu,Y., Wyman,D.,
            Ye,W.J., Zhao,J. and Zody,M.
  TITLE     Direct Submission
- JOURNAL   Submitted (17-OCT-1998) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Oct 17, 1998 this sequence version replaced gi:3738104.
            All repeats were identified using RepeatMasker: Smit, A.F.A. &
            Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html.
FEATURES             Location/Qualifiers
     source          1. .166701
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="hRPK.150_K_15"
                     /clone_lib="RPCI-11 human BAC library"
                     /map="17"
                     /chromosome="17"
     repeat_region   complement(376. .440)
                     /rpt_family="MLT1F"
     repeat_region   complement(443. .498)
                     /rpt_family="(TAA)n"
     repeat_region   complement(499. .667)
                     /rpt_family="AluSp"
     repeat_region   complement(728. .922)
                     /rpt_family="MLT1F"
     repeat_region   complement(989. .1107)
                     /rpt_family="MLT1F"
     repeat_region   2540. .2820
                     /rpt_family="AluSq"
     repeat_region   2822. .2845
                     /rpt_family="(CAAA)n"
     repeat_region   3543. .3701
                     /rpt_family="MIR"
     repeat_region   complement(3751. .4060)
                     /rpt_family="AluSx"
     repeat_region   4303. .4364
                     /rpt_family="L2"
     repeat_region   4451. .4610
                     /rpt_family="L2"
     repeat_region   complement(5380. .5541)
                     /rpt_family="MER5A"
     repeat_region   6403. .6488
                     /rpt_family="MER53"
     repeat_region   complement(6419. .6499)
                     /rpt_family="MER53"
     repeat_region   7127. .7201
                     /rpt_family="L2"
     repeat_region   complement(7696. .7734)
                     /rpt_family="MER5B"
     repeat_region   complement(7735. .8051)
                     /rpt_family="AluSx"
     repeat_region   complement(8052. .8160)
                     /rpt_family="MER5B"
     repeat_region   complement(8161. .8216)
                     /rpt_family="L1MB5"
     repeat_region   complement(8217. .8517)
                     /rpt_family="AluSx"
     repeat_region   complement(8518. .9493)
                     /rpt_family="L1MB5"
     repeat_region   complement(9500. .9542)
                     /rpt_family="MER5B"
     repeat_region   11520. .11572
                     /rpt_family="MIR"
     repeat_region   11962. .12039
                     /rpt_family="MIR"
     repeat_region   12069. .12934
                     /rpt_family="L2"
     repeat_region   complement(13277. .13326)
                     /rpt_family="(CA)n"
     repeat_region   14448. .14498
                     /rpt_family="L2"
     repeat_region   14961. .15091
                     /rpt_family="MER96"
     repeat_region   15575. .15982
                     /rpt_family="LTR16B"
     repeat_region   16458. .16738
                     /rpt_family="L1"
     repeat_region   complement(17112. .17211)
                     /rpt_family="MIR"
     repeat_region   complement(17298. .17441)
                     /rpt_family="purine-rich"
     repeat_region   18010. .18297
                     /rpt_family="AluSx"
     repeat_region   18569. .18610
                     /rpt_family="MADE1"
     repeat_region   complement(18967. .19251)
                     /rpt_family="MER7C"
     repeat_region   19252. .19556
                     /rpt_family="AluSg"
     repeat_region   complement(19557. .19860)
                     /rpt_family="MER7C"
     repeat_region   20913. .21017
                     /rpt_family="MIR"
     repeat_region   21034. .21261
                     /rpt_family="L2"
     repeat_region   21257. .21369
                     /rpt_family="MIR"
     repeat_region   21588. .21888
                     /rpt_family="AluSg"
     repeat_region   complement(22305. .22557)
                     /rpt_family="AluSx"
     repeat_region   complement(22820. .23126)
                     /rpt_family="AluSg"
     repeat_region   complement(23432. .23509)
                     /rpt_family="MIR"
     repeat_region   24275. .24536
                     /rpt_family="L2"
     repeat_region   24823. .25210
                     /rpt_family="L2"
     repeat_region   25217. .25489
                     /rpt_family="AluSx"
     repeat_region   complement(25490. .25509)
                     /rpt_family="(GA)n"
     repeat_region   complement(26510. .26597)
                     /rpt_family="MIR"
     repeat_region   28125. .28482
                     /rpt_family="L1ME1"
     repeat_region   complement(29532. .29819)
                     /rpt_family="AluY"
     repeat_region   complement(33605. .33787)
                     /rpt_family="MIR"
     repeat_region   complement(34706. .34769)
                     /rpt_family="MIR"
     repeat_region   complement(35265. .35311)
                     /rpt_family="(CA)n"
     repeat_region   complement(35478. .35498)
                     /rpt_family="(CA)n"
     repeat_region   35832. .35929
                     /rpt_family="MER5A"
     repeat_region   complement(36445. .36712)
                     /rpt_family="AluSg"
     repeat_region   36923. .36972
                     /rpt_family="MIR"
     repeat_region   36973. .37271
                     /rpt_family="AluY"
     repeat_region   37272. .37477
                     /rpt_family="MIR"
     repeat_region   complement(37478. .37524)
                     /rpt_family="(CAT)n"
     repeat_region   37635. .37821
                     /rpt_family="MIR"
```

```
     repeat_region   37895. .38191
                     /rpt_family="L1"
     repeat_region   complement(38192. .38501)
                     /rpt_family="AluSg"
     repeat_region   38502. .38768
                     /rpt_family="L1"
     repeat_region   38768. .39400
                     /rpt_family="L1MC/D"
     repeat_region   complement(39478. .39766)
                     /rpt_family="AluJo"
     repeat_region   39767. .39801
                     /rpt_family="AT_rich"
     repeat_region   complement(42375. .42461)

alignment_scores:
            Quality:    7.00              Length:      7
              Ratio:  1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-126-945B-2_COPY_236_243 x AC005747   ..

 Align seg 1/1  to: AC005747  from: 1  to: 166701

       2 AspSerGluValAspSerSer 8
         ||||||||||||||||||||
  136917 GACAGTGAGGTGGACAGCAGC 136937
```

THIS PAGE BLANK (USPTO)